

CLAIMS:

We claim:

1. A keyword generation system comprising:
 - a content parser configured to parse individual words and phrases in a selected portion of content;
 - a dictionary of words and phrases specific to a particular domain associated with said content;
 - a list of keyword candidates comprising a plurality of words and phrases specific to said particular domain;
 - a counter for each of said words and phrases in said list; and,
 - a keyword generation process both coupled to each of said content parser, dictionary, said list, and said counter and also programmed to identify said words and phrases specific to said particular domain in said selected portion of content, to write said identified words and phrases to said list of keyword candidates, to increment said counter for each of said words and phrases in said list each time said keyword generation process locates each of said words and phrases in said selected portion of content, and to select one or more of said words and phrases in said list as keywords for said content based upon said counter for each of said words and phrases in said list.
2. The system of claim 1, further comprising a list of common words coupled to said keyword generation process.
3. A keyword generation method comprising the steps of:

locating words and phrases in a selected portion of content, said words and phrases being specific to a particular domain;

adding a single instance of each of said located words and phrases to a list of keyword candidates;

for each located word and phrase which already had been added to said list of keyword candidates, incrementing a counter associated with said located word and phrase; and,

selecting keywords from said list of keyword candidates based upon words and phrases in said list having a highest counter value.

4. The method of claim 3, further comprising the step removing from consideration from said selected portion of content each of every word and phrase in said list of keyword candidates and words and phrases which are common in nature.

5. The method of claim 3, further comprising the steps of:
detecting a variation in font attributes in said selected portion of content;
selecting a string in said selected portion of content affected by said variation;
and,
adding said string to said list of keyword candidates.

6. The method of claim 5, further comprising the step of subsequently identifying said string as a word and phrase which is specific to said particular domain.

7. The method of claim 3, further comprising the step of repeated performing the locating, adding and incrementing steps for selected chunks of said selected portion of content until no content remains to be processed.

8. A machine readable storage having stored thereon a computer program for keyword generation, the computer program comprising a routine set of instructions which when executed by the machine cause the machine to perform the steps of:

locating words and phrases in a selected portion of content, said words and phrases being specific to a particular domain;

adding a single instance of each of said located words and phrases to a list of keyword candidates;

for each located word and phrase which already had been added to said list of keyword candidates, incrementing a counter associated with said located word and phrase; and,

selecting keywords from said list of keyword candidates based upon words and phrases in said list having a highest counter value.

9. The machine readable storage of claim 8, further comprising the step removing from consideration from said selected portion of content each of every word and phrase in said list of keyword candidates and words and phrases which are common in nature.

10. The machine readable storage of claim 8, further comprising the steps of:
detecting a variation in font attributes in said selected portion of content;

selecting a string in said selected portion of content affected by said variation;
and,
adding said string to said list of keyword candidates.

11. The machine readable storage of claim 10, further comprising the step of subsequently identifying said string as a word and phrase which is specific to said particular domain.

12. The machine readable storage of claim 8, further comprising the step of repeated performing the locating, adding and incrementing steps for selected chunks of said selected portion of content until no content remains to be processed.